Postal Regulatory Commission
Submitted 11/24/2020 4:28:05 PM
Filing ID: 115166
Accepted 11/24/2020

UNITED STATES OF AMERICA
POSTAL REGULATORY COMMISSION
WASHINGTON, DC 20268-0001

Before Commissioners: Robert G. Taub, Chairman;
Michael Kubayanda, Vice Chairman;
Mark Acton;
Ann C. Fisher; and
Ashley E. Poling

Periodic Reporting
(Proposal Six)

Docket No. RM2020-13

PUBLIC REPRESENTATIVE COMMENTS ON PROPOSAL SIX

November 24, 2020

I. EXECUTIVE SUMMARY

The Postal Service has submitted a mail processing variability study which differs little from previous models submitted to the Commission for review. In R2006-1 the Commission discussed four conditions the Postal Service must adequately address in any subsequent mail processing model it might submit for Commission review. With the exception of showing that data quality was good, the conditions reflected the Commission's conclusion that an appropriate model of mail processing must explicitly model the mail processing demand by distinguishing mail streams by mail shape and by outgoing and incoming sortation, which would be determined by managerial decisions to optimize efficiency as mail flowed from outgoing sort to final incoming sort.

The Commission also enumerated many other concerns it said should be addressed were another model to be submitted, including forgoing the use of productivity screens, developing data appropriate for the type of mail processing model discussed immediately above, foregoing an analysis which assumed the seperability of

machine processeses within a plant, analyzing and then deleting erroneous data at the day-level before aggregating it to higher levels.

The Postal Service has not performed any of the items requested by the Commission. For this reason, and because, the Public Representative enumerates the extent of wrongly rejected data, as well as many inconsistencies, he strongly recommends the Commission reject the model proposed by the Postal Service.

TABLE OF CONTENTS

II. INTRODUCTION

On September 15, 2020, the Postal Service filed a petition pursuant to 39 CFR 3050.11, requesting the Commission to initiate a rulemaking proceeding to consider changes to analytical principles relating to periodic reports (*Petition*). Specifically, the Postal Service presents a series of arguments supporting the timeliness of an econometric estimation of the mail processing variabilities of certain aspects of three mail sorting machines: the Delivery Bar Code Sorter (DBCS), the Advanced Flat Sorting Machine100 (AFSM100), and the Flat Sequencing System (FSS). *Petition of the United States Postal Service for the Initiation of a Proceeding to Consider Proposed Changes in Analytical Principles (Proposal Six), Submitted September 15, 2020 (Variability Study or Study*).[1]

The Postal Service first notes that "[o]ver the course of the R97-1, R2000-1, and R2006-1 rate cases, empirical mail processing variabilities based on analysis of Postal Service operating data were extensively litigated, though the Postal Rate Commission ultimately declined to adopt empirical mail processing variabilities citing an array of data quality and methodological issues." *Study, 4.* Prior to R97-1. The *Study* then states that "[c]hanges in mail volumes, Postal Service cost structure, and availability of data … merit a reassessment of the evidence on mail processing volume variability." Changes in mail volumes specifically refer to the greater percentage reduction in automated flat volumes (sorted by AFSM100 and FSS equipment) than automated letter volumes" sorted by DBCS equipment. *Id., 4-5.* The Public Representative will refer to these types of sorting equipment as "machine-types."

The *Study* also suggests that volume, runtime, and workhour data are reasonably accurate, but notably fails to show it has cured the "mountains of erroneous

[1] Docket No. RM2020-13, Petition of the United States Postal Service for the Initiation of a Proceeding to Consider Proposed Changes in Analytical Principles (Proposal Six), filed September 15, 2020. The Petition was accompanied by a study supporting its proposal. *See* also, Analysis of Labor Variability for Automated Letter and Flat Sorting, by A. Thomas Bozzo, and Tim Huegerich, (*Variability Study*). The Postal Service also filed a notice of filing of public and non-public materials relating to Proposal Two: USPS-RM2020-13/1 - Public Material Relating to Proposal Six, and USPS-RM2020-13/NP1 - Nonpublic Material Relating to Proposal Six, filed September 15, 2020.

data,"[2] identified in previous Commission Orders, and which were the primary reason the Commission rejected all previous mail processing variability studies.[3] *Id., 15*.

III. PREVIOUS COMMISSION CONCERNS

A. Introduction

In this section of his comments, the Public Representative will review the data quality and econometric concerns which led the Commission to reject mail processing variability models proposed by the Postal Service in R97-1, R2000-1, R2005-1 and R2006-1, focusing on concerns which remain valid for a model with Fixed Effects by plant and machine-type, uses total pieces fed (TPF) as the independent variable, and either runtime or workhours as the primary dependent variables of interest. These concerns will establish the issues the proposed *Study* must adequately address before the Commission adopts the proposed model, variability estimates, and its recommendation to annually update the variability estimates for these, and only these, machine-types using the proposed methodology.

B. R97-1

1. Introduction

In Docket No. R97-1, the Commission concluded that the quality of the underlying data were too poor to support a valid statistical model, that it was not based on a well-articulated economic theory, and "that the resulting variabilities (76 percent) were so low as to be implausible." *PRC Op. R97-1, Volume 1 (R97-1, Volume1), 79-96.* Specifically, the Commission identified three types of problems with the mail processing

[2] *See,* Docket No. R-2005-1, Op., Appendix I, especially para. 57.

[3] "The accuracy of workhours thus tends to be more variable than processing equipment's operating statistics, though at relatively high levels of aggregation (such as total workhours for major equipment types). USPS believes the data to be relatively accurate." *Id, 3*. "In contrast to TPF and runtime, which are machine statistics, workhours are derived from time clock rings reported to MODS through the Time and Attendance Collection System (TACS), the USPS electronic timekeeping system. The accuracy of workhours thus depends on the extent to which employees are clocked into operation codes corresponding to their actual work activities. The accuracy of workhours thus tends to be more variable than processing equipment's operating statistics, though at relatively high levels of aggregation, USPS believes the data to be relatively accurate." *Id., 15*.

variabilities proposed in this docket: model specification; inappropriate use of fixed effects; poor data quality (dirty data) and overly broad application of estimated variabilities." *Id., 65-67.*[4]

2. Model Specification Issues

a. A Cost Equation Does Not, By Itself, Suggest a Non-Arbitrary Functional Form

The Commission determined that the mail processing model the Postal Service proposed in R97-1 did not accurately characterize the relationship between volume and costs at the facility level and therefore did not appropriately specify an appropriate econometric model. Specifically, the Commission described the difference between a cost equation, which estimated a statistically significant relation between volumes and hours or cost, and a cost function, which developed an "econometric model appropriate for the set of mail processing volume levels that can be produced from a given level of inputs, the objectives postal managers pursue in choosing the inputs necessary to process the mail at the various facilities, and the inputs Postal Service managers can alter in response to a sustained increase in postal volume, among other factors." *R97-1, Volume II, Appendix F (R97-1 Appendix F), 6*.

The Commission concluded that because a cost function is not derived from an economic or econometric model, the translog model specification was not grounded in economic theory, was not necessarily better than other specifications, and therefore was open to providing arbitrarily determined variability estimates. Had other model

[4] The Commission rejected the proposed model for other reasons; e.g., it concluded the Postal Service did not address the issue that a mail processing cost driver must be proportionate to volume in order to distribute volume variable costs to sub-classes of mail. PRC Op. R97-1, Volume II, Appendix F (R97-1, Appendix F), 18.The Public Representative will not address this issue because the Commission's criticisms in this docket rest on the discrepancy between TPH and First Handled Pieces (FHP). Since the Public Representative will recommend the Commission reject the proposed model and request the Postal Service to develop the accurate FHP data so that a mail processing model along the lines of the model proposed by Mark Roberts in 2000 could be estimated without using instrumental variables to correct for mismeasured FHP data. *See, Roberts, M., An Empirical Model of Labor Demand for Mail Sorting Operations, May 31, 2002, filed at https://www.prc.gov/archived-documents?keys=mark+roberts*.

specifications been tested and found to be less robust than other functional forms, it might have been justified, but the proposed model had not been so tested.[5] *Id, 7-8.*

b. Inappropriate Use of Fixed Effects

The Commission concluded that it was inappropriate to use a fixed effects model to control for unobserved variables within plants over time. In a mail processing context this would account for unobserved heterogeneity across plants.[6] The Commission explained that the proposed model's formulation of Fixed Effects did not consider that unobserved, site specific, variables such as capital composition, facility age, degree of support costs, and space utilization, among other factors, are not exogenous, but vary with and across plants over time.[7] *See, R97-1 Appendix F, 10.*

c. Dirty Data

The Commission recognized that MODS data are often inaccurate. In general, the Commission has rejected cleaning methods which are overly broad, and do not clean data at the daily level. In R97-1, the Commission found that the Postal Service's productivity screens excluded a large percentage of sampled data without examining whether this screening method eliminated extreme values that were accurate, as well as those that were erroneous.

---

[5] In R2005-1, the Commission accepted a flexible form production function, which could be an appropriate functional form when technological or other knowledge about the underlying cost generation process, was not known. *See*, Docket No. R2005-1, USPS-T-14, Testimony Of Michael D. Bradley On Behalf Of United States Postal Service, April 8, 2005, 28. However, Bradley provided a detailed discussion of the operational features of city carrier delivery at that time, and chose variables designed to capture those operational differences. See, for example, the incorporation of delivery technology differences at page 17, and the choice of independent variables of interest according differing operational delivery constraints caused differences among mail shapes, at page 25. In any case, the functional form proposed in this proceeding is not a flexible functional form.

[6] "Under a strict exogeneity assumption on the [unobserved] explanatory variables, the fixed effects estimator is unbiased: roughly the idiosyncratic error, *u*, should be [non-arbitrarily] uncorrelated with each [time-demeaned] explanatory variable across all time periods." See, Wooldridge, J., Introductory Econometrics, 3rd ed. (*Woolrdige*), 2006, 486.

[7] "However, variables such as the degree of support costs, space utilization, degree of flex labor, as well as several others, vary over time for the same facility and are persistently different across facilities. If it is important to control for these differences in facilities over time in recovering the relationship between mail processing costs and mail volume, then witness Bradley's fixed-effects estimation procedure is unable to yield a valid estimate of this relationship." *Ibid.*

> "The estimated equations are incapable of accurately representing behavior because all of the data needed to reveal responses under the most extreme circumstances has been systematically removed." *Id., 27.*

The Commission found it unacceptable that the productivity filters, as well as two other filters the Postal Service used in this proceeding, had deleted approximately 12.9 percent of pieces sorted on DBCS machines, and 15.7 percent of pieces sorted on automated flat sorting machines (FSM and other automated flat sorting machines). *Id., 30.* The Commission also found using a screen which removed, without examination, (as little as) one percent of the observations from each tail of the productivity distribution was unacceptably high.

> "The only persuasive evidence provided…that the productivity scrub removes data subject to data entry error is…testimony that some of the reported productivities are plainly beyond the capacity of the machines and personnel at the site. The Commission shares witness Neels's suspicion that much of the deleted data is merely unusual and not erroneous. It is the Commission's understanding that deleting observations solely because they are unusual is not considered good econometric practice [because it]…is very possible that such 'unusual' observations contain the most information about the true relationship between cost and volume. *Id., 33.*

C. R2000-1

The PR will not provide a broad identification of the issues discussed in R2000-1, because they either duplicated the issues discussed in R97-1, or because they primarily dealt with issues related to proper measure of sub-class distribution keys.[8] However, the Commission did identify one new issue related to "Dirty Data," which it later expanded upon in R2005-1, namely the problem that data which is erroneous at the daily level, may appear to be non-problematic when aggregated to a monthly or quarter level. "MODS data are actually derived from reports collected by the Postal Service at a

[8] The Public Representative will not discuss this issue, since his he will recommend the Commission reject the instant proposal for not having adequately addressed the issues the Commission enumerated in R2006-1. He will not recommend an alternate model or recommend the Commission adopt alternate variability estimates, which if accepted, would require distributing modified attributable machine costs to different products.

much finer level of detail and were first aggregated even before being given to either witness. UPS witness Neels has described in testimony how aggregation masks reporting errors leaving them undetectable by the screens*." Docket No. R2000-1, Volume II, Appendix F, 40.* The Commission concluded that "filters" or "screens" which removed obvious errors, would not remove erroneous daily errors aggregated to the monthly level. Moreover, there was no evidence which showed whether or not the "hidden" and remaining erroneous observations were systematic or random. *Id., 41.* The Commission concluded that this problem created an unaddressed errors-in-variables problem of unknown magnitude. *Id., 44.*

D. R2005-1

1. Introduction.

The Postal Service modified certain aspects of its mail processing variability model in this docket. In particular, it used an "instrumental variables" approach to reduce downward bias due to erroneous observations. *Docket No. R2005-1, Op., 77, Chapter IV, November 11, 2016, 77.* The Commission did not accept the proposed model, partly because the rate changes proposed in the docket were the result of a negotiated settlement by affected parties, and partly because the mail processing data remained unfit for econometric estimation. *Id., 78*. The Commission did perform an extensive analysis on these data, which had not been adequately addressed on the record which in previous mail processing dockets. The Commission requested the Postal Service to address the data and econometric problems which had been identified since R97-1 in the next general rate case. *Ibid.*

3. Further Exploration of Data Quality

a. Aggregation of Daily Errors Prevents Correction and Results in Biased Estimates

The Commission extensively discussed the ways in which errors which are observable at the daily level can lead to unobserved errors when aggregated to the monthly or quarterly level, and the type(s) of econometric problems which could be result. MODS data are collected daily, but have been aggregated to the monthly level

when presented to a mail processing econometrician, without first attempting to remove gross errors or to distinguish missing observations from zero-valued observations. Incorrect or missing observations are combined with correct observations in the aggregated data sets. The Commission concluded that errors left in the aggregated data sets can result in biased and inconsistent variability estimates. *Docket No. R2005-1, Op., Appendix I (R2005-1 Appendix I), November 11, 2016, 1.*

b. Mismatched Observations

The Commission illustrated how erroneous daily observations could become mismatched, but appear reasonable, when aggregated to the monthly or quarterly level. It illustrated how two or more observations with erroneous time and volume values might combine to create the appearance of a valid data point at a more aggregated level. A simple example could occur if one daily observation had zero hours and 100,000 pieces recorded, and an observation from another day had 50,000 hours, and zero pieces recorded. When rolled up to the monthly level, the monthly observation would appear as 100,000 pieces and 50,000 hours. Observations with zero valued time but positive volume, or zero valued volume and positive time would be detectable at the daily level, but would "disappear" at the monthly level. The combined piece-hour observation might appear within the range of acceptable piece-hours and would not be captured by a one or 5 or 10 percent screen. However, the true values of the two observations might be very different, still within a range that would not be captured by a screen, but (if enough of them existed) could yield a different variability estimate. This problem would be categorized as a mismeasurement of both dependent and independent variables, the latter which could create an "errors-in-variables" problem.

E. R2006-1

1. Introduction

Significant insights into the controversies associated with the estimation of mail processing variabilities were introduced and, to an extent, resolved in R2006-1 by the testimonies submitted by the Postal Service, the Office of Consumer Advocate, and UPS. The Commission summarized the controversies and the advances made in

R2006-1 in Appendix I of its Opinion and Recommended Decision.[9] The Public Representative will summarize the Commission's findings with respect to the advances made in this proceeding. These include a theory of demand for mail processing, choice of an appropriate functional form, and measurement issues raised by the choice of appropriate variables.

4. The Commission Rejected the Postal Service's Model and Results

The Commission has repeatedly criticized previous Postal Service's mail processing models for not having developed a theory of mail processing demand which would "advise" an analyst about the most reasonable types of variables, their measurement, and the relations among the variables – in other words an econometric model. In this proceeding, the Commission approved of the Postal Service's recognition of data quality problems, but criticized its continued failure to remove erroneous data at the daily level. *Docket No. R2006-1, Opinion and Recommended Decision (Op.), Volume I, February 26, 2007, 31-34*. It also rejected the Postal Service's method of modelling each sorting process separately, unrelated to other mail processing operations.[10,11] *Ibid.*

5. The Commission Accepted The Basic Features of OCA's Mail Processing Model Submitted By Mark Roberts in R2006-1

The Office of Consumer Advocate (OCA) submitted a mail processing model by Mark Roberts,[12] which the Commission generally endorsed "as making substantial

[9] See, respectively, Docket No. R2006-1, USPS-T-12, Direct Testimony of A. Thomas Bozzo on Behalf of the United States Postal Service, filed May 3, 2006 , Office of Consumer Advocate, OAC-T-1,Direct Testimony of Mark Roberts, On behalf of the Office Of The Consumer Advocate, filed September 9, 2006, and Op. Postal Rate Commission, Volume I and Volume II, February 26, 2007.

[10] "The Postal Service models, do not capture the impact of volume on the mix of operations in the plant, and therefore, measure only a narrow category of volume variability." *Ibid.*

[11] A good deal of the Commission's discussion in this Opinion is devoted to the appropriateness of using instrumental variables to correct for poor quality or mismeasured data, in particular, FHP. Since, the proposed model does not use Instrumental Variables, the PR will not discuss this issue.

[12] Direct Testimony of Mark J. Roberts on Behalf of the Office of the Consumer Advocate the Postal Rate Commission, OCA-T-1, (*Roberts*) October 26, 2006.

headway toward the goal of estimating the marginal cost of mail processing labor." *Id., 21*. Roberts' model of mail processing defines output as real volume, i.e., the number of *unique* pieces of mail processed in the plant, which was best measured by the volume of First Handled Pieces (FHP).[13] (emphasis added)

In his theory of production there are separate sorting processes (sets of machines and manual processes) for each major mail shape—letters, flats, and parcels, and that managers respond to changes in volume levels during the day by adjusting the mix of sorting technology (both automated and manual) used within a shape stream (letters, flats, parcels) in order to minimize costs or meet service obligations. *Roberts, 3.* Since the flow of each mail shape through a plant is a mix of unsorted and presorted originating mail and the destinating mail stream is a mix of mail sorted to different final levels, he treats the plant as producing two different outputs for each mail shape: the number of letters, flats, parcels receiving an initial (outgoing) sort, and a final (destinating) sort. *Id., 4.*[14]

All parties used Instrumental Variables to correct for error-ridden volumes and hours in MODS. The Commission rejected the use of Instrument Variables for this purpose and concluded that "… econometric modeling of mail processing labor demand variability seems to have reached an impasse," *Op., Volume I, 48,* unless the Postal Service developed a comprehensive substitute measure of volume which measures RPW volumes exiting each mail processing plant. *Id., 44.*

6. Commission Requirements For Accepting Subsequent Mail Processing Variability Studies

The Commission laid down five requirements future mail processing studies must meet in order to be seriously considered: a) they must be based on a theory of mail

[13] According to the most recent MODs Manual, "FHP records mail volume in the operation where it receives its first distribution handling. A first handling piece is a letter, flat, or parcel that receives its initial distribution in a Postal Service facility. Each mailpiece distributed in an office receives one and only one FHP count." *See*, Management Operating Data System (MODS), Handbook M-32 September 2018, 11,

[14]

processing demand which distinguishes mail streams by mail shape and by outgoing and incoming sortation; b) they must allow and account for the substitution among processes used by a particular mail shape during its mail flow from outgoing to incoming final sortation, including manual operations; c) they must be based on valid data on the capital stock by operation, properly matched to the time of deployment and retirement; d) they must use daily (perhaps tour-level) data based on a measure of volume which measures RPW volumes exiting each mail processing plant, e) they must use valid instruments if measurement error is a concern. This last requirement would involve using location-specific information that exists with respect to processing plants in the network, and measurements of local economic activity, and local business and consumer profiles expected to drive mail volume, *Id., 36-49.*

IV. POSTAL SERVICE VARIABILITY MODEL

A. Model Description

1. Justification Of A New Variability Model

The Postal Service justifies proposing another model of mail processing variability by noting that it has proposed many mail processing models, beginning with Docket No. R97-1, but the Postal Rate Commission did not accept any of them "citing an array of data quality and methodological issues." *Variability Study, 4*. It justifies proposing another mail processing study, because there have been "changes in mail volumes…cost structure, and data availability" since R2006-1. *Ibid.*

It notes that there has been a significant decline of demand for letters and flats since a local peak in FY 2006. Letter volumes declined by 31 percent and Flat volumes declined by 44 percent. *Ibid*. The report does not argue that volume declines alone merit reexamination of mail processing variabilities. Rather, it appears to argue that the greater decline in flats, along with a greater increase in unit costs, is the primary, volume-related justification.[15] *Id., 5.*

[15] "While the aggregate outcome of "real" sorting costs declining roughly in proportion to volumes may seem relatively benign among USPS's challenges, the aggregates mask significant differences for

The Report notes that "technological factors" may also justify submitting another mail processing variability study.[16] One the one hand, letter sorting technology has kept the staffing level required to support its operation relatively constant, at one mail processing clerk per machine. In contrast, many AFSM100's now use "Automatic Induction" (AI), which are essentially mail preparation machines. They are staffed by three clerks who prepare flats so they may be rapidly fed into the AFSM100 by the one clerk involved in the direct sorting operation. The report states that "'integer constraints' may limit the limit downward flexibility of labor usage when volumes are low and/or declining." *Id., 6*. This could mean that there is limited ability of the clerks involved in incidental allied preparation operations, from either clocking out of their operation if the flow of mail to be inducted falls below full utilization level, or it may mean the Postal Service is constrained from consolidating AFSM100 or FSS machines to reduce the number of mail processing clerks and the hours they incur.[17]

Later, the Postal changes the characterization of automated flat sorting machines, from machines with a relatively fixed complement of clerks engaged in incidental allied preparation operations, to sets of differently automated sorting machines, by vintage.

---

the letter- and flat-shape mailstreams as well as for component products. Volumes of flat-shape mail fell markedly faster than letter mail—46 percent versus 31 percent—and upward pressure on unit costs for flat-shape products from sorting operations along with cost coverage issues for some flat-shape mail products has led to increased Commission scrutiny of Postal Service flats costs and operations…." *Ibid.*

[16] While the Commission's Initial Order states that the Postal Service has "submitted a new methodology for estimating volume variabilities for certain mail processing cost pools," the Public Representative will argue the Postal Service's methodology differs little from the one it relied upon in R2006-1, except for its assumption that data quality are reliable and so it did not use Instrumental Variables to correct for data deficiencies. Docket No. RM2020-13, Public Regulatory Commission (PRC) Notice Of Proposed Rulemaking On Analytical Principles Used In Periodic Reporting (Proposal Six), September 23, 2020, 2. He will show that the Postal Service does not address most of the requirements the Commission requested the Postal Service address before submitting another mail processing variability study.

[17] Later, the Report maintains that the dearth of consolidation of all three automated sorting machines has accounted for a decline in economies of scale since FY 2015. *See, Figure 5, Id., 13.* This should mean that the faster decline in AFSM and FSS economies of scale should not be related to the pace of consolidation since FY 2015.

> "flat sorters including the AFSM 100 and FSS are variable throughput machines and, ***depending on the details of machine configurations***, may also be operated with *variable* complements of feeders, sweepers, and prep workers. The staffing index and/or throughput…may vary positively or negatively with processing workloads. For instance, increasing the staffing index can, to some extent, increase throughput towards a machine's technical limits at some cost to productivity. If the staffing index and/or machine throughput depend on the TPF volume processed on the equipment, even the runtime portion of labor demand for the machines will not necessarily be unit-elastic." *Id., 7-8*. (emphasis added)

The limitations a clerk or manager has with regard to responding to an unexpected increase in volume is more likely to refer to older vintage AFSM100s, than AFSM100s with Automatic Induction. The Report also repeats arguments the Postal Service has previously made, and the Commission rejected, namely that letter and flat sorting machines include set-up and take-down time – activities such as printing container labels, positioning trays, and sweeping bins. The Report argues (again) that these activities do not vary directly with volume. *Ibid.*

7. Model Specification
    a. Data Used To Inform Model Specification

The Report makes six conclusions after examining monthly MODS and TACS data (which are matched to create a time-volume observation):

- The accuracy of workhours depends on the extent to which clerks are clocked into the proper mail processing operations. Consequently, although workhours vary more than runtime, the Postal Service believes aggregating workhours from the daily to the monthly level, makes workhours relatively accurate. *Id., 15.*
- Within the sample regression period (FY2016 – FY2019) workhours varied by month for all three automated machines. Monthly or Seasonal variation of workhours increased in the following order, from lowest to highest monthly variation: AFSM100, DBCS, and FSS; *Id., 17-19.*

- Technological characteristics, management decisions and different network roles played by different plants were assumed to vary across plants independently of volume;[18] *Id., 20*.
- The Model expects workhours to vary by annual managerial decisions, but less so due to the previous month's volume, or runtime in general. *Ibid.*
- There was substantial consolidation, technological change (primarily for AFSMs and FSS) and network operating changes between FY2007 to FY2015, while these factors were relatively constant between FY2016 and FY2019. *Id., 21*.
- Elasticities using workhours as the dependent variable were somewhat sensitive to the inclusion of outliers with unusual values for labor productivity. *Ibid.*
- Observations with extreme productivity values below or above 5 percent tails were considered to be caused by random or idiosyncratic factors. *Ibid.*

b. Model Specification

The Report chose to use a log-linear functional form. While a log-log functional form can be the basis for a flexible form production or cost function, it would not permit dummy "month" variables to be used, since all observations would be deleted when the logarithm of a dummy value of zero were to be taken. The finding that volume and workhours vary by month, led to the inclusion of monthly dummy variables to account for peak periods. The finding that technological characteristics vary across plants independently of volume led to using a FE model with machine clusters to account for unobserved effects of capital composition within a plant over time.

The finding that workhours varied differently than runtime informed the choice to run two basic regressions: one with the independent variable constructed as the log of runtime and the other with the independent variable constructed as the log of workhours. Estimations using each dependent variable option showed that when extreme values of productivity, the observations with the five percent largest and

[18] "…the intercepts a_i *potentially* depend on volume-independent technological parameters, management considerations affecting staffing levels locally, and specific plants' processing network roles. (emphasis added), *Ibid.*

smallest productivity values (by machine type over the sample period) were deleted, variability estimates improved.

The logarithm of Total Pieces Fed (TPF) was chosen as the primary independent variable, because "TPF is a direct machine count of the articles inducted by the machine for automated sorting operations, including both successfully sorted items as well as machine rejects, and is the primary driver of machine runtime as well as a driver of other labor requirements." *Ibid., 12.*

Finally, because staffing at a plant in any particular month may be subject to short term constraints, the model includes a one month lag of volume as an independent variable. Furthermore, because Postal Service management often makes decisions using year-to-year or Same-Place-Last-Year (SPLY) comparisons, the model includes an annual lag on volume to capture any possible annual changes made by management which might affect volume and/or workhours. *Id., 20.*

Algebraically, the regression of runtime (RT) and then workhours (H) on TPF may be expressed as follows:

$$\ln(RT) = a_i + b_1 \ln(TPF_{it}) + b_2 \ln(TPF_{i,t-1}) + b_3 \ln(TPF_{i,t-12}) + cDM_t + \varepsilon_{it} \quad (1)$$
$$\ln(H) = a_i + b_1 \ln(TPF_{it}) + b_2 \ln(TPF_{i,t-1}) + b_3 \ln(TPF_{i,t-12}) + cDM_t + \varepsilon_{it} \quad (2)$$

In equations (1) and (2), cDM_t represents a linear combination of monthly dummy variables. The elasticity is calculated as the sum of coefficients on the logarithmically transformed volume variables -- $b_1, b_2, and\ b_3$.

8. Regression Results

Table 1 below summarizes the results of equation (1), and Table 2 summarizes the results of equation (2).

Table 1
Regression Results for Runtime Models,
FY2016-FY2019 Sample Period

Operation	Elasticity	F-Statisic	N
DBCS	0.958	2,1685*	9,062

AFSM100	0.771	1.365*	7,973
FSS	0.60	529*	1,763

Asterisk indicates F-value is statistically significant.

Table 2
Regression Results for Workhour Models,
FY2016-FY2019 Sample Period

Operation	Elasticity	F-Statistic	N
DBCS	0.976	830*	9,062
AFSM100	0.774	221*	7,973
FSS	0.804	223*	1,763

Asterisk indicates F-value is statistically significant.

The Postal Service chose results from the "Hours" equation (2), with current TFP, monthly lagged TFP, annually lagged TPF, and monthly dummy variables were individually and jointly significant. Table 3 represents the results of an Hours regression without lagged TPF variables or monthly dummies.

Table 3
Regression Results for Workhour Models With No Lagged TPF Variables,
FY2016-FY2019 Sample Period

Operation	Elasticity	F-Statistic	N
DBCS	.805	46.0*	10,074
AFSM100	.749	89.8*	8,059
FSS	.593	399.5*	1,792

Asterisk indicates F-value is statistically significant. Note: The F-statistics Reported in Table 3 have not been transformed to account for the Fixed Effect Regressions. *See*, USPS-RM2020-13-1, Analysis, analysis.txt.

Although the Hours regression without lagged TPF variables or monthly dummies had significant elasticity estimates for each machine type, the Postal Service

presumable did not choose them because more significant information was contained in the regression which produced Table 2, and F-statistics were substantially larger. Id., 24. However, the Postal Service did not explicitly defend its preference of a lagged model with monthly dummies. Nor, did it explain why it rejected the "Runtime" presented in Table 1.[19]

Finally, the Postal Service investigated the reasonableness of the elasticities within the overall FY2007-FY2019 period, by running rolling regression analyses using the log-linear models, albeit using elasticity estimates that did not include the lagged TPF variables. *Variability Study, 24.* It reports stable elasticity estimates during the sample period of FY2016 to FY2019 for DBCS, and AFSM (but not FSS) based on a 48 and 60 month rolling average. *Id., 26.*

V. ANALYIS OF POSTAL SERVICE'S VARIABILITY MODEL

A. The Postal Service Has Not Met The Burden Of Proof Necessary To Submit A New Model Of Mail Processing Variability

As presented in Section III.E.4 above, the Commission discussed four conditions the Postal Service must adequately address in any subsequent mail processing model it might submit for Commission review. With the exception of showing that data quality was good, the conditions reflected the Commission's conclusion that an appropriate model of mail processing must explicitly model the mail processing demand by distinguishing mail streams by mail shape and by outgoing and incoming sortation,

[19] Presumably the Postal Service rejected the Runtime Model because fewer monthly dummy coefficients were individually significant for the AFSM100, and they were not jointly significant, and the monthly lag of TPF for FSS was not individually significant, although the lagged and non-lagged volume variables were jointly significant, as were the monthly dummy variables. *See*, USPS-RM2020-13-1, Analysis, analysis_lag_seasonal_tests.txt.

which would be determined by managerial decisions to optimize efficiency as mail flowed from outgoing sort to final incoming sort.

This would require using a measure of volume where mail entering the plant would be equal to mail exiting the plant. At the beginning of R2006-1, the use of First Handled Pieces (FHP), with instrumental variables to correct for measurement error, seemed to be a promising approach, rather than modeling each type of sorting process (both manual and mechanized) as completely separable one from the other.

The Commission ultimately rejected the ability of the instrumental variables used to correct for measurement errors in FHP. Rather, than reject the above-described model, the Commission supported this approach, and expected the Postal Service to either develop a comprehensive substitute measure of volume which measures RPW volumes exiting each mail processing plant. *Id., 44.* Needless to say the Postal Service, not only did not develop the data necessary to implement the Commission's proposed model, it did not make any attempt to address the Commission's desires.

If future recommendations by the Commission are to have any meaning, in light of the complete failure of parties to even address its recommendations, it must reject the Postal Service's proposed mail processing variability model.[20]

B. The Postal Service Did Not Address The Commission's Concerns About Hidden Errors Due To Aggregation.

The section on R2006-1 also illustrated the Commission's concern that the Potal Service could only identify and delete truly erroneous data at the most disaggregated level it was available; certainly not when aggregated to the monthly level. The Commission consistently rejected the Postal Service's continual asides that aggregated

[20] The Commission expected creating acceptable data would be a formidable task, which is why it concluded by saying that "For more than a decade, the Commission has expressed concern that the quality of the MODS data upon which mail processing variability models depend is too poor to support valid econometric models. Over that time, the quality of that data has not improved in any discernable respect. This record has clarified the theoretical requirements of valid mail processing models. At the same time it illustrates that marginal costs in much of its mail processing network were greater than 100 percent. The Postal Service should understand that unless the quality of the MODS data improves, or alternative data is developed, models that rely on such data have little prospect of being accepted by the Commission. R2006-1, Op., Volume I, 52.

data would wash out serious anomalies. At times aggregation can capture erroneously recorded data at a more disaggregated level, but whether this occurs depends on whether the erroneous variables of interest are properly specified at the more aggregated level.

An example from the City Carrier Street Time Model will illustrate this point. The record shows detailed discussion concerning the manner in which managers would "break off" parts of an individual carriers' volume on his or her daily delivery route, and assign it to a carrier assigned to a different route, but within the same ZIP Code. Both volume and associated hours would be recorded at the same daily ZIP Code level, even though delivered (think processed) to different routes. Because the data remained at the daily ZIP level, managers and analysts could examine the transferred volume and time to determine whether either volume or time were measured with error. This is not the case with the aggregation performed in the Postal Service's proposed variability model. The erroneous daily hours and/or volume cannot be identified at the aggregated monthly level.

C. The Technological Reason(s) The Report Uses To Explain Why Flat's Variabilities Are Substantially Less Than 100 Percent Cannot Be Distinguished From Exogenous Management Decisions

The Variability Report offers numerous explanations supporting the notion that mail processing variability for automated flat's sorting machines have volume variabilities less than 1. The Public Representative has already noted that one reason – the flat sorting technology features constant throughput and fixed staffing rates, subject to an integer constraint, which limits managements' ability to adjust workhours to varying volumes, conflicts with its later explanation that the limitations a clerk or manager has with regard to responding to an unexpected increase in volume is more likely to refer to older vintage AFSM100s, than AFSM100s with Automatic Induction than throughput limitations on advanced AFSMs and FSSs subject to integer constraints. *See first, Id., 6, then Id., 7-8.*

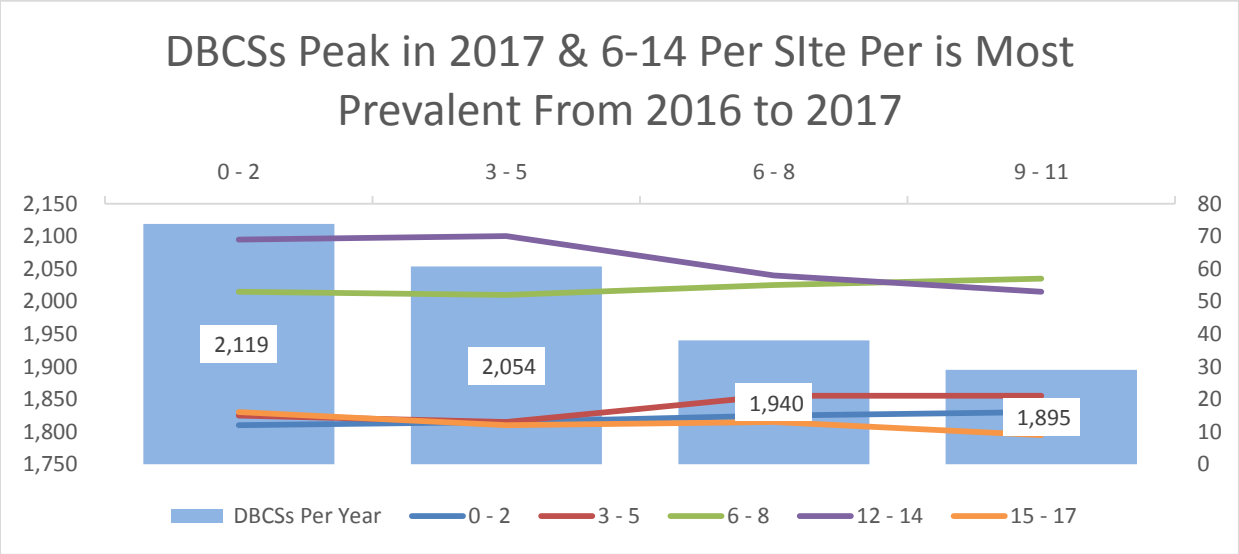
The Postal Service admits there may be no certain way to distinguish its technological determinist approach from the Commission focus on USPS's

management of flats operations. *Id., 5.* After analyzing machine volumes per plant over time, the Public Representative is inclined to support the Commission's view that management decisions to either consolidate or eliminate equipment is the primary reason for relative low flat's productivities and elasticities (estimated at the machine level) less than one.

The PR modified the Postal Service program which created the final analysis dataset, by stopping the program before the "reshape" command was given. This allowed the data to be placed in excel with the Machine-Type Variable (Group2) explicitly identified and matched with volume and time data.[21] Since, the data included number of each type of machine at each plant, the PR examined the distribution of each machine type, by number of machines per plant over the sample period. He concluded that consolidation of machines per plant was very limited for DBCS, AFSM, and FSS. Rather the primary managerial decision was the much greater reduction of DBCSs compared to AFSMs and in particular, FSSs. Charts 1-3 illustrate the limited change in the distribution of number of machines (by machine type) from FY2016 to FY2017, while showing notable decline in machine count.

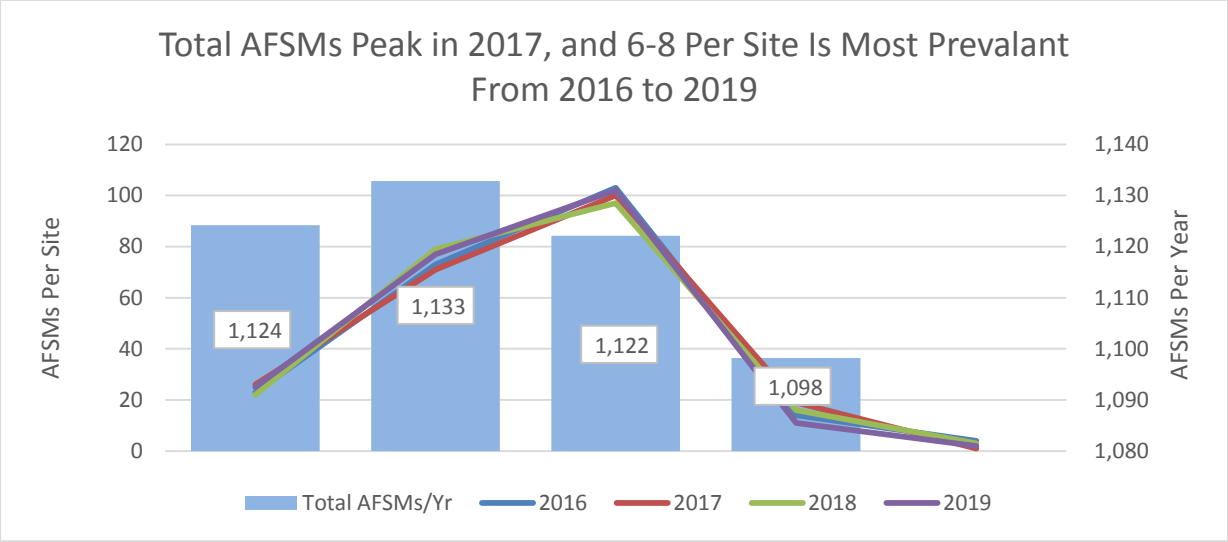
Chart 1

[21] See, PR-LR-1, pr_analysis_set.do, pr_analysis_set.dta, import stata export xls.sas, and pr_anal_grp pivot.xlsx.



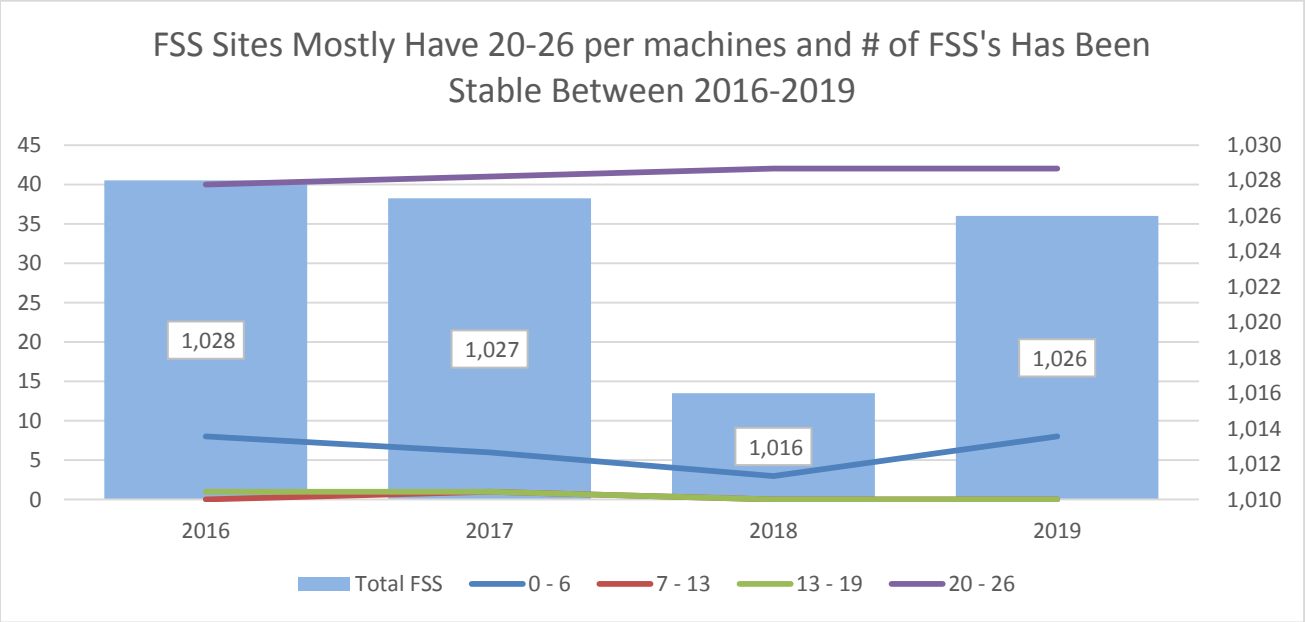
Source: PR-LR1-p pr_anal_grp.xlsx

Chart 2



Source: PR-LR1-p pr_anal_grp.xlsx

Chart 3



Source: PR-LR1-p pr_anal_grp.xlsx

Charts 1 – 3 presume that if the number of DBCS, AFSM100, and FSS machines per plant have approximately the same distribution over the sample period, it means that there has been very little, if any consolidation of machines from plants with many (or few) machines to plants with fewer (or more) machines. However, Chart 1 shows that the number of DBCS machines authorized to be remain in plants by Postal Service Managers fell from 2,119 FY2016 to 1,895 in FY2019, a decline of 224 or a 10.6 percentage point decline. Chart 2 shows a similar stable distribution of AFSM machines per plant over time, but the number of AFSM machines fell from 1,124 in FY2016 to 1,098 in FY2019, a decline of 26 or a 2 percentage point decline authorized by Postal Service management. Chart 3 shows, except for a one year decline, a very stable number of FSS machines per plant, and a decline of FSS machines from 1,028 in Fy2016 to 1,026 in FY2019, a decline of 2, or a 0.2 percentage point decline, also a management decision. The charts show that Postal Service managers have matched declining letter volumes, discussed by the Report with a substantial decline in letter processing machinery, and the attendant operating and labor costs, while management has failed to match declining flat volumes with a matching decline in flat processing

machinery. In the case of letters, managers kept the decline in volumes in sync with processing costs, by reducing machinery, thereby keeping the elasticity stable and close to one. In contrast, managers have not kept the decline of flats volumes in sync with processing costs by reducing the stock of AFSM100s and FSSs. The result has been a lower and declining variability over time.

D. Much Of The Postal Service's Analysis Is Inconsistent Or Results-Driven

1. The Report's Productivity Screens Are Results-Driven

The Postal Service has once again relied on broad productivity screens, which filter out extreme values, without considering whether they are truly erroneous. The Postal Service also chose to use a 5 percent productivity screen because it produced results which looked better, namely, a few more of the coefficients for dummy variables were significant using the 5 percent rather than 1 percent screen. The Postal Service states this results-drive decision in somewhat cryptic terms: "[a]s noted in the response to part (a), the 5th/95th percentile criterion was intended to balance the desires to exclude erroneous data and to maintain samples with high coverage of the population of costs under study. *Responses Of The United States Postal Service To Questions 1-11 Of Chairman's Information Request No. 1, Response to question 6.a, October 14, 2020.* It's hard to take this response seriously. The 5 percent filter retains fewer observations than the one percent filter. Consequently, the answer must mean that the 5 percent filter did a better job filtering out erroneous data. The Postal Service does not show that the 5 percent filter does a better job filtering out erroneous observations. It only shows that it filtered out more observations. Yet, a greater percentage of these observations could be valid than if the one percent filter had been applied.[22]

The Public Representative examined the data screened out of each of the three machine samples using the one, five and 10 percent filters. He performed a scatter plot of observations to determine truly extreme values, and then calculated how many non-

[22] Recall that the Commission criticized the Postal Service's practice of automatically filtering out extreme values, without knowing whether the all of the deleted data were in fact problematic, or were even outliers which affected the regression results.

zero observations had been deleted which were inside the acceptable range of each scatter plot. Table 4 summarizes his findings.

Table 4
Percent of Observations Wrongly Deleted By 1%, 5%, and 10% Screens
By Machine Type

	Productivity Screen		
	One Percent	Five Percent	Ten Percent
AFSM	22.7%	95.4%	97.7%
DBCS	16.5%	96.7%	98.3%
FSS	13.1%	98.7%	98.7%

Source: PR-LR-1, Tholds and Deleted Obs.xlsx

The evidence shows that it was a gross mistake to choose the 5 percent productivity screen, and was due to the failure to show any interest in examining the data deleted by the screens. This is a bit surprising, since the Commission has criticized the Postal Service every time it has used this practice as part of its many, proposed mail processing variability studies. It's hard not to conclude that the 5 percent screen was chosen because it obtained the best results, since there is little difference in "coverage" between the 5 and 10 percent screens.

2. Another Inconsistency Is The Choice To Use TPF/Hour As The Productivity Screen For The Runtime Equation

The Postal Service considered two basic regressions, after having seen that including these variables improved results. Good econometric practice would expect the analyst to enumerate the model which follows from his or her economic theory of the situation to be analyzed, develop the appropriate variables, and perform the analysis. Only if problems arise, should addition of variables be considered.

The Postal Service develops two possible equations: one which regresses runtime against volumes, and one which regresses hours against volumes. Since the both models assume each sorting operation is separable and independent from all other operations which occur in the plant, the appropriate model is one which would regress runtime against volume. Instead, the Postal Service chose the model which regressed

hours against volume because it achieved more results which were statistically significant across both explanatory and control variables. *See, Id., 21-23.*

3. Allied Time Should Have Been Removed From Flat's Regressions

Another inconstancy is the inclusion of incidental allied labor time in the AFSM100 and FSS regressions, but not in the DBCS regressions. The Postal Service should have removed flat preparation operations from the flat's regressions in order to produce a consistent set of elasticity estimates to compare.

4. It Was Inconsistent And Results-Driven To Use TPH/Hour As The Productivity Screen For The Runtime Regressions

Similarly, the Postal Service failed to match productivity screen definition with the type of regression being run. The Postal Service noted that "estimated elasticities for workhours were somewhat sensitive to the inclusion of outliers with unusual values for labor productivity." *Id., 21.* The measure of productivity it used for both the runtime and hours regressions was TPF/Workhours. In order to be consistent, the productivity screens for the runtime regression should have been defined as TPF/Runtime.

5. There Is Nearly Perfect Correlation Between Actual Volume and Lagged Volume Independent Variables.

The Postal Service's decided to include lagged volumes variables with little justification other than that there might be some additional information included in model results.[23] Id., 20. In response to a Commission Request for Information about potential collinearity among the current and lagged volume variables, the Postal Service provided tests which showed nearly perfect collinearity among current and lagged volume variables for each of the three machine types. *See, USPS-FY2020-13-3, analysis_seasonal_chir2vif.txt.* The Postal Service justifies including variables which are

[23] While there is relatively little reason to expect that machine runtime should materially depend on workloads other than current-period TPF, workhours may have a longer adjustment process due to limitations on the flexibility of USPS labor. The inclusion of lagged TPF terms allows for adjustment processes of workhours with respect to workloads over longer time scales. Notably, same period last year (SPLY) reporting is a common piece of management information provided by various USPS data systems, including MODS, for managing workhours. Staffing may also be subject to shorter-term constraints. The extended models include the first and twelfth lags of monthly TPF, the latter being the same month in the previous year."

highly collinear and have very high VIF values, because a joint test of significance among the volume variables showed they were jointly significant. *Ibid.* However, a test of joint significance of variables which are highly collinear almost guaranteed to find the variables are jointly significant. This is because regression procedures cannot capture the shared effects or information. Kennedy, P., A Guide To Econometrics, 6E, 2008, 194. Since the relevant information is shared among the collinear variables, it stands to reason that a test which jointly includes their effects would show them to be jointly significant.

For the reasons enumerated in the above discussion, the Public Representative requests the Commission to adopt his recommendations.

Respectfully Submitted,

Lawrence Fenster
Public Representative

901 New York Avenue, N.W., Suite 200
Washington, DC 20268-0001
Phone (202) 789-6862
Email: larry.fenster@prc.gov